P5_nadolsky_0710, ANL-HEP-CP-01-105, FERMILAB-Conf-01/330-T, MSUHEP-11025, SMU-HEP 01/11

PDF uncertainties in WH production at Tevatron

P.M. Nadolsky[1,2*] and Z. Sullivan[3,4†]
[1] *Department of Physics & Astronomy, Michigan State University, East Lansing, MI, 48824*
[2] *Department of Physics, Southern Methodist University, Dallas, TX, 75275*
[3] *High Energy Physics Division, Argonne National Laboratory, Argonne, IL, 60439*
[4] *Theoretical Physics Department, Fermi National Accelerator Laboratory, Batavia, IL, 60510-0500*

(Dated: 29 October, 2001)

We apply a method proposed by members of CTEQ Collaboration to estimate the uncertainty in associated W-Higgs boson production at Run II of the Tevatron due to our imprecise knowledge of parton distribution functions. We find that the PDF uncertainties for the signal and background rates are of the order 3%. The PDF uncertainties for the important statistical quantities (significance of the Higgs boson discovery, accuracy of the measurement of the WH cross section) are smaller (1.5%) due to the strong correlation of the signal and background.

arXiv:hep-ph/0110378 v2 21 Nov 2001

The steady improvement of world hadronic data has stimulated significant interest in quantitative estimates of theoretical uncertainties due to incomplete knowledge of parton distribution functions (PDFs) [1-7]. The published studies discuss the impact of uncertainties on the shape of the PDFs or simple observables, such as the total cross section for the $W-$boson production at Tevatron. It is interesting to apply the proposed methods to more involved observables, such as cross section distributions or the ratio of the signal to background in the search for new physics.

To illustrate new issues that such application involves, consider how PDF uncertainties affect the potential of discovery of the Standard Model Higgs boson with the mass $M_H = 115$ GeV via associated WH production at Tevatron. In this process, Higgs bosons can be discovered by observing excess production of $b\bar{b}$ pairs with the invariant mass close to M_H, *e.g.,* in the band $95 \leq M_{b\bar{b}} \leq 135$ GeV. The WH signal consists of two tagged b-quark jets, a lepton, and missing transverse energy associated with the unobserved neutrino from the W decay. The dominant background process is direct $Wb\bar{b}$ production. In an experimental analysis, this QCD background would be estimated by extrapolation of the $M_{b\bar{b}}$ distribution from the regions of $M_{b\bar{b}}$ where the WH cross section is negligibly small, *e.g.*, from the side bands $75 \leq M_{b\bar{b}} \leq 95$ GeV and $135 \leq M_{b\bar{b}} \leq 155$ GeV.

We model both the signal and background with tree-level matrix element calculations using MADGRAPH [8] at a hard scattering scale $\mu^2 = \hat{s}$. To simulate the resolution of the hadron calorimeter, we smear the jet energies with a Gaussian of width $\Delta E_j/E_j = 0.80/\sqrt{E_j} \oplus 0.05$ (added in quadrature). We simulate the acceptance of the detector by using the selections listed in Table I. An isolation cut is placed on the lepton, as defined by a cone of radius ΔR. We use the impact-parameter b-tagging efficiency function defined in SHW 2.3 [9].

The PDF uncertainties influence the potential for the discovery of the Higgs boson in several ways. First, they affect the shape of the $M_{b\bar{b}}$ distribution and, therefore, the accuracy of the extrapolation of the background from the side bands. Second, the relative errors for statistical quantities, such as the ratio S/B of the signal and background rates, may differ significantly from the relative errors for S and B if the latter ones are correlated or anticorrelated. Third, the errors for S and B are likely to be different in the positive and negative directions, commonly due to the changes in the shape of the distributions $d\sigma/dM_{b\bar{b}}$ under the variation of the PDFs. Hence, the integrated distribution may be less constrained in the positive direction than in the negative direction.

As a first step in the study of the above issues, we estimate the PDF uncertainties for S and B with the method proposed by J. Pumplin, D. Stump, Wu-Ki Tung *et al.* (PST) [7]. The PST method is based on diagonalization of the matrix of second derivatives for χ^2 (Hessian matrix) near the minimum of χ^2. Since χ^2 is approximately parabolic near its minimum χ_0^2, hypersurfaces of constant χ^2 are hyperellipses in the space of the original 16 PDF parameters $\{a_i\}$. By an appropriate change of coordinates $\{a_i\} \rightarrow \{z_i\}$, $i = 1, \ldots, 16$, we can transform hyperellipses into hyperspheres. We assume that all acceptable PDF sets correspond to χ^2 that does not exceed its minimal value χ_0^2 more than by T^2. As a result, the acceptable PDF sets have $\{z_i\}$ within a sphere of the radius T^2 around $\{z_i(\chi_0^2)\} \equiv \{z_i^0\}$. We present the results for $T = 10$. Our global analysis of the PDFs uses the same set of hadronic data as in Ref. [7].

Table I: Cuts used to simulate the acceptance of the detector at the Tevatron run II.

$\lvert\eta_b\rvert < 2$	$E_{Tb} > 20$ GeV
$\lvert\eta_l\rvert < 1.5$	$E_{Tl} > 20$ GeV
$\lvert\Delta R_{bl}\rvert > 0.7$	$\not{E}_T > 20$ GeV

*Electronic address: nadolsky@pa.msu.edu; work supported by the National Science Foundation, U. S. Department of Energy, and Lightner-Sams Foundation
†Electronic address: zack@fnal.gov; work at Argonne supported by the U. S. Department of Energy under contract No. W-31-109-ENG-38; work at Fermilab supported by the U. S. Department of Energy under contract No. DE-AC02-76CH03000.

The PDF uncertainty for an observable O is the maximal change in O as a function of variables $\{z_i\}$ varying within the tolerance hypersphere. The PST method estimates the variation of O as

$$\delta O = \sqrt{\sum_{i=1}^{16} \delta O_i^2}, \text{ where } \delta O_i \equiv T\frac{\partial O}{\partial z_i} \approx T\frac{O(z_i^0 + t) - O(z_i^0 - t)}{2t}, \tag{1}$$

and $t = 5$ is a small step in the space of z_i. For brevity $O(z_1^0, \ldots, z_i^0 \pm t, \ldots, z_{16}^0)$ is denoted as $O(z_i^0 \pm t)$.

The PDF error (1) is a combination of 32 cross sections, each of which is known with some uncertainty due to the Monte-Carlo integration. One might be concerned about accumulation of Monte-Carlo errors in the process of calculation of δO. Fortunately, this accumulation does not happen, because the calculation of δO in Eq. (1) and the propagation of Monte-Carlo errors involves only summation in quadrature. The Monte-Carlo error $\Delta_{MC}\delta O$ for δO is given by

$$\Delta_{MC}\delta O = \frac{1}{\delta O}\left(\frac{T}{2t}\right)^2 \sqrt{\sum_{i=1}^{16} \left(O_i - O_{i+1}\right)^2 \left(\Delta_{MC}O_i^2 + \Delta_{MC}O_{i+1}^2\right)}, \tag{2}$$

where $\Delta_{MC}O_i$ are Monte-Carlo errors for O calculated with the PDF set i. If all $\Delta_{MC}O_i$ are approximately the same ($\Delta_{MC}O_i \approx \Delta$ for $i = 1, \ldots 32$), Eq. (2) simplifies to

$$\Delta_{MC}\delta O \approx \frac{T}{2t}\Delta\sqrt{2}. \tag{3}$$

Hence, in this case the Monte-Carlo error for δO is proportional to the Monte-Carlo error for O_i and not to the number of the PDF parameters. In our calculation, the Monte-Carlo uncertainty of δO does not exceed 20% of δO.

In addition to the symmetric error δO, it is useful to estimate maximal variations of O in the positive and negative directions, given by

$$\delta O_- = \frac{T}{t}\sqrt{\sum_{i=1}^{16} \left[\max\left(O(z_i^0) - O(z_i^0 + t), O(z_i^0) - O(z_i^0 - t), 0\right)\right]^2}, \tag{4}$$

$$\delta O_+ = \frac{T}{t}\sqrt{\sum_{i=1}^{16} \left[\max\left(O(z_i^0 + t) - O(z_i^0), O(z_i^0 - t) - O(z_i^0), 0\right)\right]^2}. \tag{5}$$

These variations define the true allowed range for O and may differ significantly from δO. For instance, δO for the total cross section of $W^{\pm}$-boson at the LHC is 5%, while δO_- and δO_+ are -3 and $+7$%, respectively.

In the associated WH production at the integrated luminosity $\int Ldt = 15$ fb^{-1}, we expect the following numbers of the signal and background events in the signal band $95 \leq M_{b\bar{b}} \leq 135$ GeV:

$$S = 49.7^{+1.8}_{-1.4}\,(3.0\%), \quad B = 110.1^{+3.6}_{-4}\,(3.1\%). \tag{6}$$

The number in parentheses is the symmetric relative error for S or B estimated with the help of Eq. (1). Common statistical combinations of S and B are

$$S/B = 0.451^{+0.011}_{-0.006}\,(1.5\%), \qquad S/\sqrt{B} = 4.73^{+0.13}_{-0.07}(1.8\%), \tag{7}$$

$$\sqrt{S+B}/S = 0.254^{+0.004}_{-0.006}(1.7\%). \tag{8}$$

According to Eqs. (7-8), the errors for the statistical quantities are very asymmetric. In the lower ($75 \leq M_{b\bar{b}} \leq 95$ GeV) and upper ($135 \leq M_{b\bar{b}} \leq 155$ GeV) side bands, we expect $94.5^{+2.3}_{-5.7}(4\%)$ and $30^{+3.3}_{-0.06}(5\%)$ background events, respectively.

Eqs. (6-8) also show that the PDF errors on S/B, $S/\sqrt{B}$, $\sqrt{S+B}/S$ are smaller (1.5%-1.8%) than the uncertainties on S and B ($\sim$ 3%), which signals a correlation between the PDF errors for S and B. We can get a feeling of this correlation by studying correlations of individual variations $\delta S_i/S$ and $\delta B_i/B$ (Fig. 1). On average, the magnitude of $\delta S_i/S$ is larger than the magnitude of corresponding $\delta B_i/B$, since the "average ratio" of these magnitudes $\frac{1}{16}\sum_{i=1}^{16}|(\delta S_i/S)/(\delta B_i/B)| = 1.53$ exceeds unity. However, we are more interested in the correlation of the largest values of $\delta S_i/S$ and $\delta B_i/B$, which give dominant contributions to the total relative errors of S, B, and statistical quantities. According to Fig. 1, the largest $\delta S_i/S$ and $\delta B_i/B$ are well correlated,

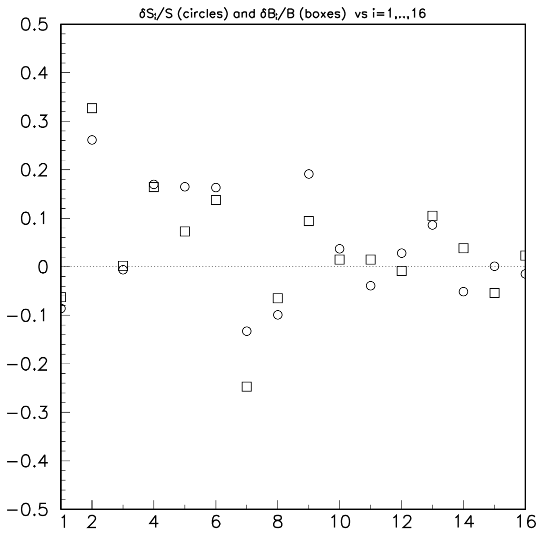

Figure 1: Spread of $\delta S_i/S$ (circles) and $\delta B_i/B$ (boxes) for $i = 1, \ldots, 16$.

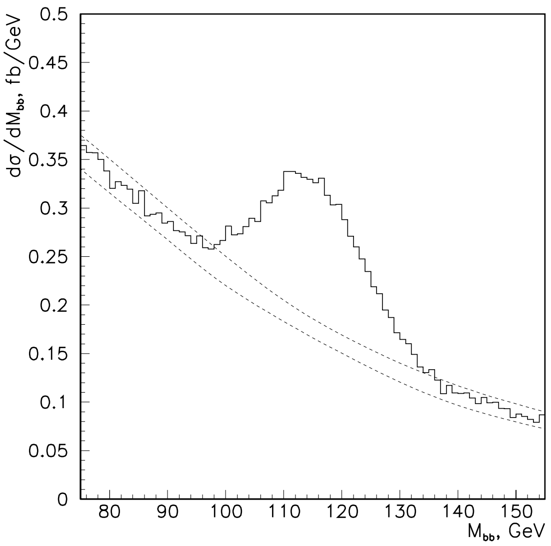

Figure 2: The total signal plus background $M_{b\bar{b}}$ distribution (solid line) as compared to the PDF uncertainty band for the background (dashed lines).

so that their contributions to $\delta(S/B)/(S/B)$, etc., cancel. As a result, the relative errors for the statistical quantities are smaller than the relative errors for S and B.

The overall correlation of vectors $\{\delta S_i/S\}$ and $\{\delta B_i/B\}$ can be quantified by introducing the cosine of the angle between these vectors [7]:

$$\cos\varphi \equiv \frac{1}{\delta S\,\delta B}\sum_{i=1}^{16}\delta S_i \delta B_i . \tag{9}$$

Then, the relative error for $A \equiv S/B^p$ is

$$\left(\frac{\delta A}{A}\right)^2 = \left(\frac{\delta S}{S}\right)^2 + \left(\frac{\delta B^p}{B^p}\right)^2 - 2\frac{\delta S}{S}\frac{\delta B^p}{B^p}\cos\varphi, \tag{10}$$

and correlated or anticorrelated S and B correspond to $\cos\varphi = 1$ or -1, respectively. Similar correlation angles can be calculated for any pair of relative errors, including the errors for backgrounds in the upper and lower side bands. We find that $\cos\varphi$ for S and B, $S+B$ and S in the signal band are 0.89, 0.95, respectively, *i.e.*, the correlation is very good. The correlation cosine between the background cross sections in the lower and upper side bands is also large (0.62), which indicates that the PDF uncertainty mostly affects the overall normalization of the $M_{b\bar{b}}$ distribution and not its shape. Correspondingly, the extrapolation from the side bands accurately approximates the background in the signal band. Fig. 2 illustrates the relative size of PDF uncertainties for the background in comparison to the signal plus background distribution.

To conclude, we propose to use asymmetric PDF errors and correlations between PDF errors in detailed studies of PDF uncertainties. Using these quantities, we find that the cumulative effect of the PDF uncertainties on the significance for the discovery of the Higgs bosons at Tevatron is not large ($\sim 1.5-1.8\%$). The PDF uncertainties for the signal and background ($\sim 3\%$) are much smaller than the eventual statistical errors for the measurement of the WH cross section ($\gtrsim 25\%$) even if Tevatron Run II accumulates 15 fb^{-1} of the integrated luminosity.

[1] S. Alekhin, Eur. Phys. J. **C10**, 395 (1999), arXiv:hep-ph/9611213.
[2] S. I. Alekhin, Phys. Rev. **D63**, 094022 (2001), arXiv:hep-ph/0011002.
[3] W. T. Giele and S. Keller, Phys. Rev. **D58**, 094023 (1998), arXiv:hep-ph/9803393.
[4] V. Barone, C. Pascaud, and F. Zomer, Eur. Phys. J. **C12**, 243 (2000), arXiv:hep-ph/9907512.
[5] M. Botje, Eur. Phys. J. **C14**, 285 (2000), arXiv:hep-ph/9912439.
[6] D. Stump et al. (2001), arXiv:hep-ph/0101051.
[7] J. Pumplin et al. (2001), arXiv:hep-ph/0101032.
[8] T. Stelzer and W. F. Long, Comput. Phys. Commun. **81**, 357 (1994), arXiv:hep-ph/9401258.
[9] M. Carena et al. (2000), arXiv:hep-ph/0010338.